Lesson 12

Computerizing Public Health Surveillance Systems

Instructor's Guide Form

Lesson Title: Computerizing public health surveillance systems

Lesson Goals: For each learner, to be able to discuss the present situation with regards to computerizing public health surveillance systems and the future possibilities.

Lesson Objectives: By the end of this lesson, the learner will be able to:

1) describe the technical possibilities in computerization of surveillance

2) explain the gap between what is possible and the systems of today

3) list the barriers to optimal use of computers in surveillance

4) describe the issues involved in current computerization of surveillance

5) discuss the key issues remaining to be resolved

Equipment and Materials Needed:

- Overhead projector
- Transparencies #12.1 - #12.29

Time Required: 60 minutes

Synopsis of Lesson: This lesson explores the current situation and the future possibilities in computerization of surveillance.

Return to Surveillance in a Suitcase-Table of Contents CD 98': Main Menu

Lesson 12

Computerizing Public Health Surveillance Systems

Instructor's Guide Form (continued)

Adult Education Application: At the core of most surveillance systems are individuals who collect the data and place the data on forms. Sometimes the forms are constructed in a way that facilitates the accurate collection of the data while others are not. The instructor could involve the learners in a discussion that asks them to list the characteristics of good versus bad forms. Using the learners' responses, the instructor could make two lists on an overhead or flip chart, one for the good characteristics and one for the bad characteristics. Once the learners have identified all the characteristics that come to mind, the instructor may want to add a few additional ones to assure the completeness of the list.

Lesson 12

Computerizing Public Health Surveillance Systems

Topical Outline

I. An ideal surveillance system
- A. An epidemiologist's system of the future
- B. Future surveillance

II. Barriers to the ideal surveillance system
- A. Medical-record systems
- B. Political and administrative barriers
- C. Lack of recognition that information is useful
- D. Feeling that data base must be "clean"

III. Technology of the future
- A. High-capacity storage devices
- B. Networks
- C. New user interfaces
- D. New programming tools
- E. Higher-capacity processors and more memory
- F. Video and computer integration
- G. Voice and pen input

Lesson 12

Computerizing Public Health Surveillance Systems

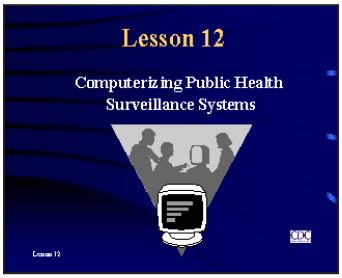
Topical Outline (continued)

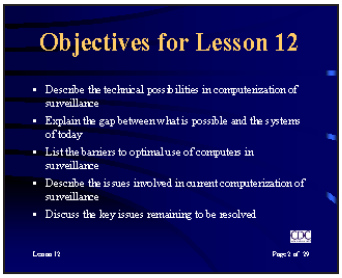
IV. Computerized public health surveillance today

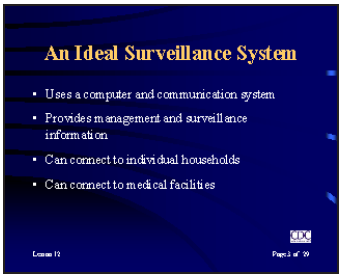
A. Morbidity and Mortality Weekly Review (MMWR) is available on-line, and can be received through electronic mail.
B. Epi Info
C. Computerizing a surveillance system
D. Basic needs
E. Files, records, files
F. Hardware
G. Software
H. Designing entry forms
I. Data entry
J. Cleaning and editing the data
K. Analysis of data
L. Distributed data base
M. Transmitting data
N. Correcting and updating records from another site
O. Individual and summary records
P. Linking special-purpose records to the main data base
Q. Preparation of report
R. Dissemination of data
S. Data disasters
T. Back-up methods
U. Training of staff and transition techniques

V. Ongoing issues

A. Standardization vs. customization and local control
B. Confidentiality vs. free access to medical records
C. Cost of information
D. Dealing with preliminary and incomplete data







Lesson 12

Computerizing Public Health Surveillance Systems

Content Outline

Lesson Objectives:

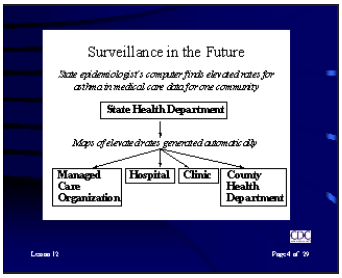
- **Describe the technical possibilities in computerization of surveillance**
- **Explain the gap between what is possible and the systems of today**
- **List the barriers to optimal use of computers in surveillance**
- **Describe the issues involved in current computerization of surveillance**
- **Discuss the key issues remaining to be resolved**

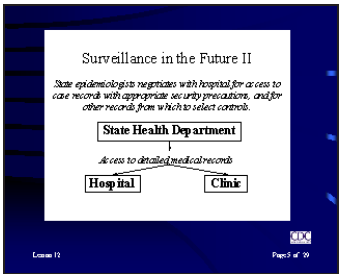
I. An ideal surveillance system

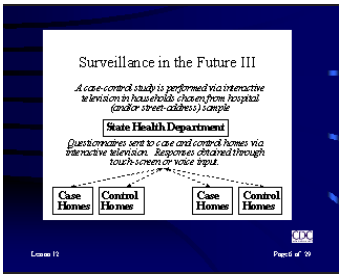
A. An epidemiologist's system of the future

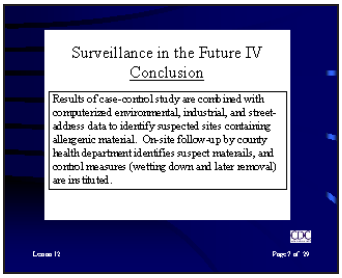
1. will have a computer and communications system
2. capable of providing management and surveillance information
3. system also capable of being connected to individual households
4. capable of being connected to medical facilities to obtain additional information

Return to Lesson 12







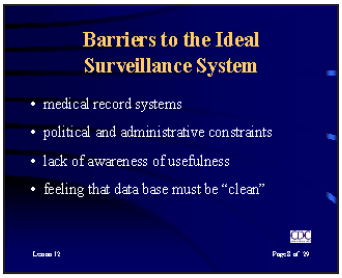


B. Future surveillance

1. responsive interactive system
2. system with automatic input from all inpatient and outpatient medical facilities
3. uses standard records for each office visit or hospital admission
4. use computer to produce a series of maps for all conditions with unusual patterns
5. can extract desired information
6. can contact households immediately using fax-like equipment
7. receive data quicker, respond to situations quicker
8. surveillance in the future: example
    a. state epidemiologist's computer finds elevated rates for asthma in medical care data for one community
    b. state epidemiologist negotiates with hospital for access to case records with appropriate security precautions, and for other records from which to select controls.
    c. case-control study is performed via interactive television in households chosen from hospital (and / or street-address) sample
    d. questionnaires sent to case and control homes via interactive television. Responses obtained through touch-screen or voice input.

Return to Lesson 12

e. Conclusion: results of case-control study are combined with computerized environmental, industrial, and street-address data to identify suspected sites containing allergenic material. On-site follow-up by county health department identifies suspect materials, and control measures (wetting down and later removal) are instituted.



II. Barriers to the ideal surveillance system

A. Medical-record systems

1. not all are computerized

2. different variables used

3. different formats

4. needs prior to development of a responsive interactive system
    a. simple core public health record
        1) age
        2) gender
        3) geographic location
        4) diagnosis
    b. needs to be available in a standard format in real time

B. Political and administrative barriers

1. need to establish a will for a national computerized medical record system

Return to Lesson 12

2. technical hurdles
    a. standard but suitably flexible record formats
    b. solutions to problems associated with confidentiality
    c. incentives to create the records (including the assurance of appropriate and cost-effective use of the records)
    d. voice input

C. Lack of recognition that information is useful

1. need to increase level of awareness that information is useful in public health and should be available to public health agencies

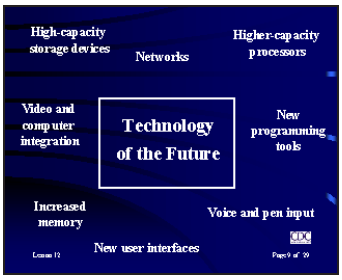
2. need to publicize that technical solutions to problems of confidentiality are available
    a. one-way encoding algorithm
    b. provide solutions to matching and follow-up problems

D. Feeling that data base must be "cleaned up"

1. is a pervasive feeling among those in charge of data that a data base must be "cleaned up" before anyone else can use it.

2. quality control is necessary and important but:
    a. the concept of surveillance includes rapid turnaround
    b. a realization on the part of everyone concerned (even the media and the public) that the data are preliminary
    c. the understanding that in order to look at today's data today, one must be willing to accept today's imperfections

Return to Lesson 12



III. Technology of the future

A. High-capacity storage devices

1. CD ROM
    a. compact disks with read only memory
    b. provide access to large bibliographic data bases anywhere there is electricity

2. MEDLARS
    a. database of the U.S. National Library of Medicine
    b. can be searched from anywhere with appropriate computer technology

3. past data bases from the U.S. and elsewhere will become available on CD ROM
    a. process of cleaning them up often reveals gaps and inconsistencies
    b. reflects changing definitions
    c. diminish value as consistent anchors for comparison

B. Networks

1. LAN
    a. local area network
    b. system linking microcomputers, terminals, and workstations with each other and a mainframe computer to facilitate sharing of equipment, programs, data, and information

2. benefits of LANs
    a. decrease in written memoranda
    b. many users can enter data in multiple computers connected by a LAN
3. costs of LANs

Return to Lesson 12

a. expense of installation and support
b. requires special software to protect against errors
c. special precautions to protect confidentiality if several people have access

C. New user interfaces

1. parts of programs that interact with users have become easier to understand and more attractive
    a. pull-down menus
    b. windows
    c. pointing devices such as the mouse

2. trend is toward a standard set of screen controls

3. improvements have costs in terms of requirements
    a. for faster computers
    b. for more memory
    c. for greater skill to produce programs

D. New programming tools

1. software still requires hundreds of thousands of lines of handwritten and highly personal "coding"

2. new trends have made programming more productive
    a. fourth-generation data bases
    b. computer-assisted software design (CASE) tools
    c. subject-oriented design

Return to Lesson 12

E. Higher-capacity processors and more memory

1. have removed many of the limits that required use of mainframes

2. now, almost any project can be done on a microcomputer or several microcomputers connected by a LAN

F. Video and computer integration

1. photographs and fully functional video

2. increases opportunities to use color and 3-D dynamic displays for epidemiologic data

3. there are possibilities for computer interaction via ordinary television sets
    a. could interview citizens via cable television
    b. results can be captured immediately in computerized form

4. new challenge in identifying responses

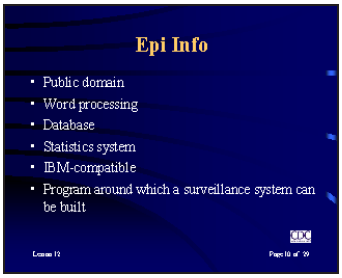
G. Voice and pen input

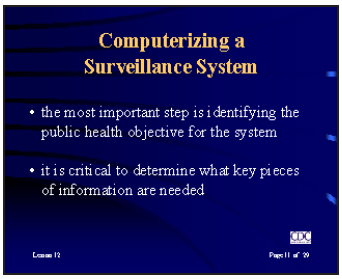
1. systems can now identify thousands of spoken words
    a. expensive
    b. allows for a crude interaction between voice and computer

2. computers that recognize handwritten text of reasonable structure are available

3. a future possibility is medical handwriting being replaced by voice dictation into a lapel microphone

Return to Lesson 12





IV. Computerized public health surveillance today

A. Morbidity and Mortality Weekly Review (MMWR) is available on-line, and can be received through electronic mail.

B. Epi Info

1. public-domain word-processing, database, and statistics package for IBM-compatible microcomputers

2. one type of database and statistics program around which a surveillance system can be built

3. developed through a joint project between CDC and the Global Programme on AIDS, World Health Organization

4. makes possible the participation of all 50 states in the National Electronic Telecommunications Surveillance System (NETSS)

5. evaluation indicated benefits
    a. improved access to data
    b. improvement in quality of data
    c. improvement in access associated with decentralized entry of data

C. Computerizing a surveillance system

1. difficult task

2. success depends on administrative and epidemiologic environment and on the software

3. most important step is identifying the public health objective for the system

Return to Lesson 12

   a. may be related to the disease occurrence
   b. may be outcome related

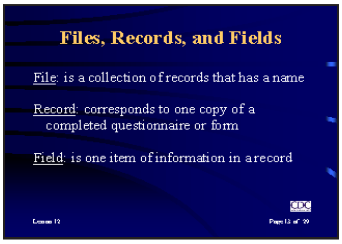
4. most successful computer systems are those that change methods by which an agency operates rather than those that merely automate a manual task

5. question to address
   a. what key pieces of information do I want to see on my desk (computer screen) every day, week, month, or year that will make my work easier or more effective?
   b. can be asked at all levels of management

6. why computerize?
   a. increase speed of processing
   b. provide graphic capability
   c. enhance analytic capability
   d. improve quality of data or of reports
   e. improve quantity of data

7. there are no systematic studies on the benefits of computerizing public health surveillance systems but some articles describe methods of evaluation
8. problems associated with developing a computerized surveillance system
   a. few commercial developers feel that it is financially worthwhile to develop software for this market alone
   b. applications such as spreadsheets, languages, and word processors may sell millions of copies to the general public, making general products more attractive to commercial developers than epidemiological software.

Return to Lesson 12



D. Basic needs

1. first requisite for computerization is a paper system or operational design that works reasonably well or would do so if the process were speedier and more accurate

2. chaos computerized is not necessarily an improvement

3. can evolve the computer facilities in small stages with minimal investment until the system proves to be useful and well-conceived

4. personnel who will handle collection of data, data entry, analysis, and system maintenance are important contributors to computerization plan

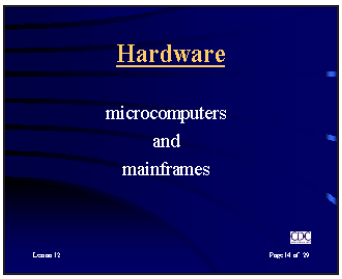
E. Files, records, and fields

1. files
    a. computerized records are stored in files
    b. a file is a collection of records, usually one record per case, that has a name
    c. files can be opened, closed, read, written to, or discarded
    d. are stored on nonvolatile media such as hard or floppy disks or magnetic tape

2. records
    a. correspond to one copy of a completed questionnaire or form
    b. usually, one disease report or questionnaire is stored in a file as a single record
    c. many records can be stored in each file

Return to Lesson 12

3. fields (examples from Epi Info system)
    a. a field is one item of information in a record, for example:
        1) NAME,
        2) AGE, or
        3) DATEONSET
    b. records in a particular file all have the same fields
    c. each field has the following components
        1) name
        2) type
            a) text
            b) uppercase text
            c) numeric
            d) date
        3) length, e.g.,
            a) 22 characters for NAME
            b) 3 characters for AGE
    d. during analysis, fields may be called variables
    e. commands such as TABLES DISEASE COUNTY are used to instruct the system to process a particular file, and construct a table



F. Hardware

1. processing time tends to reflect the record length and the number of records

2. size of each record should be kept short if large numbers will be processed

3. systems that require processing of millions of records can be reduced by sampling to a manageable size for the microcomputer

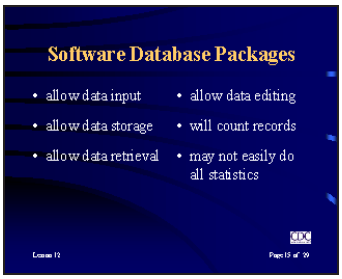
Return to Lesson 12

4. mainframe can be used to select a sample of records
    a. particular age groups
    b. every tenth record, or other methods
    c. samples can then be exported for processing on a microcomputer

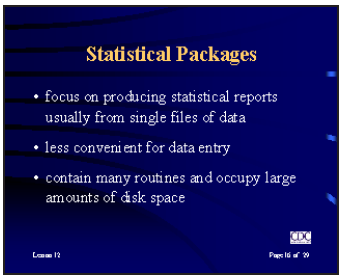
G. Software

1. often less crucial than the skills of those who will program and run software

2. the indispensable programmer problem
    a. a single expert programmer writes a system in his or her favorite language and then leaves the job
    b. beware of this



3. database packages
    a. allow data input
    b. allow data storage
    c. allow data retrieval
    d. allow data editing
    e. will count records
    f. do not easily do such statistics as odds ratios
    g. examples: dBase, Paradox, Foxbase, Clipper

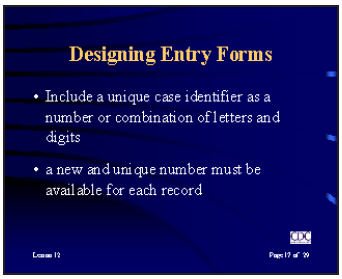


4. statistics packages
    a. focus on producing statistical reports usually from single files of data
    b. less convenient for data entry
    c. contain many routines and occupy large amounts of disk space
    d. examples: Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS)

Return to Lesson 12

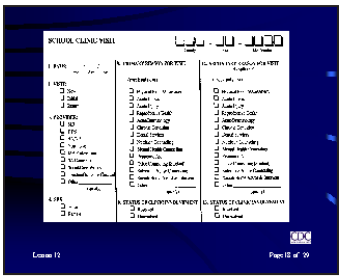
5. Epi Info
    a. fits on three 1.4 megabyte diskettes
    b. provides a combination of database and statistical functions
    c. allows relational linking of several files during data entry or analysis
    d. graphics can be produced
    e. commonly used epidemiologic statistics are part of the statistical output
    f. limitations
        1) limits on numbers of records that can be sorted or relationally linked at one time (tens of thousands)
        2) text fields are limited to 80 characters
        3) would not be a good choice if large amounts of free text are to be stored
6. limitations of software packages
    a. most statistical and database packages cost at least several hundred dollars
    b. some limit the number of fields in a record or the number of records in a file
    c. few databases will do statistics without advanced programming
    d. statistics packages may have limitations in handling textual (alpha) data
    e. most statistics packages allow processing of only one file at a time
    f. a complete surveillance system requires functions of both database and statistical packages, in addition to mapping and graphing

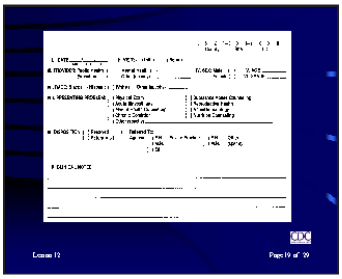
H. Designing entry forms



1. data items
    a. usually entered in a standard format
    b. information stored in files containing one record per individual

Return to Lesson 12





2. Epi Info format
    a. format is specified by typing a questionnaire or form in the word processor
    b. result resembles a paper form
    c. entry blanks are indicated by special symbols
    d. computer reads the form and constructs a file in the proper format

3. form design
    a. include a unique case identifier as a number or combination of letters and digits
    b. a new and unique number must be available for each record

4. amount of data entry and computer storage required may be minimized by computerizing only information that will actually be used

5. *Exercise: facilitate group discussion and critique of forms*

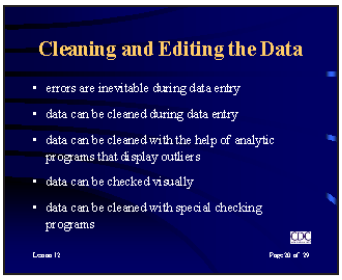
I. Data entry

1. who determines details of data entry
    a. those who prepare the paper records for entry
    b. those who enter the data

2. most disease reports are entered in batches

3. records linked to more extensive specialized forms can be sent as partial submission and revised later to avoid delays in reporting caused by the slower progress of data collection for more detailed forms

Return to Lesson 12



J. Cleaning and editing the data

1. errors or duplications inevitably occur during data entry

2. additional information may require changes or additions

3. data can be cleaned during data entry or with the help of analytic programs that display outliers

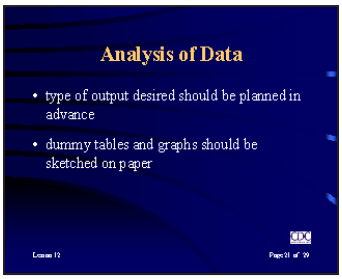
4. data can be checked visually

5. Epi Info allows extensive programming of error checks on data entry
    a. fields can be set to accept only specified codes
    b. multiple fields can be checked for inconsistencies

6. special checking programs may be used

7. errors should be caught and corrected near the time of data entry if possible

8. when to check data
    a. depends on orientation and number of personnel available
    b. try different methods to determine preferences

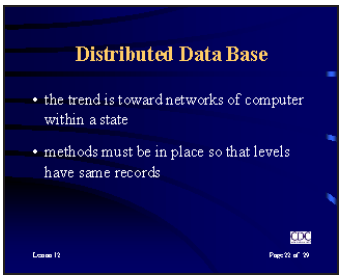


K. Analysis of data

1. type of output desired should be planned in advance

2. dummy tables and graphs should be sketched on paper
3. Epi Info and many other programs can be programmed to print a table or mixture of text and tables in almost any format (report generator)

Return to Lesson 12

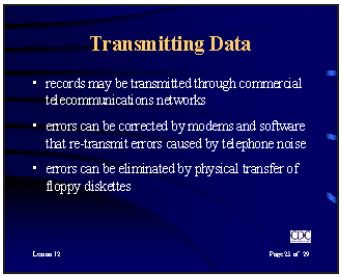


L. Distributed database

1. trend is toward networks of computers within a state
    a. connected by modem
    b. analogous to ways used in the National Electronic Telecommunications Surveillance System (NETSS)
        1) has more than 50 state and territorial participants
        2) participating site sends data periodically to a computer at the next level up

2. methods must be in place to see that both levels have the same records

3. state summaries can be compared with the contents of the national data base and discrepancies can be reported



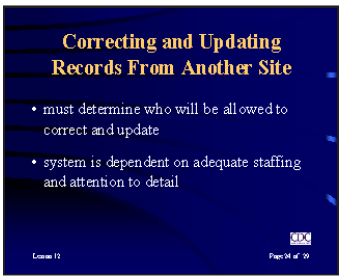
M. Transmitting data

1. in NETSS, most states transmit reports each week through a commercial telecommunications network
    a. joined together in a single file for processing on the CDC mainframe computer
    b. error checking is done to test for invalid codes and other problems
    c. error notices sent back to states

2. errors can be corrected by use of modems and software that retransmits errors caused by telephone noise

3. errors can be eliminated by physical transfer of floppy diskettes
    a. allows large files to be transferred with minimal inconvenience
    b. may be appropriate if the additional trouble of setting up modems or software is not warranted

Return to Lesson 12



N. Correcting and updating records from another site

1. in NETSS, only state participants are allowed to update records

2. updates are sent as records with the same identification number as that for the original record in the data base

3. system is dependent on adequate staffing and attention to detail

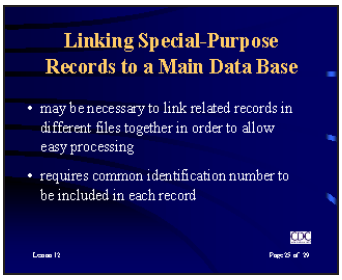
O. Individual and summary records

1. many systems function with a record for each individual case report

2. some systems have a need for summary reports
    a. helpful if large numbers of similar records are processed
    b. helpful if only summary numbers are available
    c. allows records from entire years to be summarized in condensed format

3. summary report
    a. similar to a case record
    b. contains an additional field
    c. commands sum contents rather than counting individual records

Return to Lesson 12

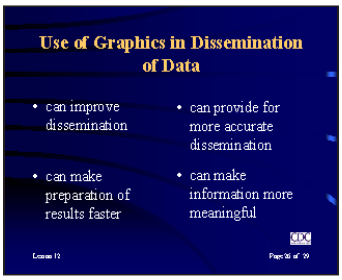


P. Linking special-purpose records to the main data base

1. may be necessary to link related records in different files together in order to allow easy processing
2. requires common identification number to be included in each record
3. Epi Info and other database programs allow automatic linking of records through a common identifier

Q. Preparation of a report

1. should meet needs of users
2. should summarize reported data
3. should use figures, tables, and charts
4. should include interpretations of analyzed data
5. should include narrative related to control and prevention
6. should include summary of public health reports from other sources



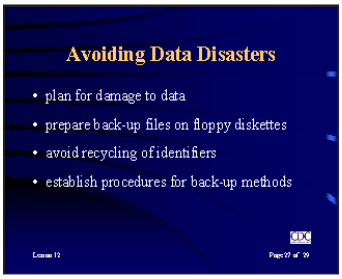
R. Dissemination of data

1. computerization can assist by making new methods of analysis or presentation practical
2. use of tabular or graphics software in conjunction with desktop publishing technology improve dissemination
    a. can make preparation of results faster
    b. more accurate
    c. meaningful

Return to Lesson 12

3. Epi Map
    a. public domain companion to Epi Info
    b. makes mapping available to anyone with an IBM-compatible

4. graphics may be made available on-line or by distributing floppy or CD-ROM disks



S. Data disasters

1. destruction or damage of data on hard disks should be expected and planned for

2. need to prepare backup files on floppy diskettes

3. avoid recycling of identification so that new records don't overwrite old ones

4. causes of problems
    a. upgrading of hardware or software
    b. computer viruses

T. Backup methods

1. backup copies should be rotated so that several circulate in turn and so that the one overwritten has at least two more recent relatives

2. protection from fire, water damage, damage by panic-stricken personnel
    a. keep at least one backup in a site remote from the computer
    b. set the write-protection feature on the diskettes after making the backup

Return to Lesson 12

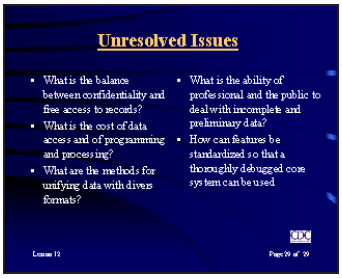
3. when to upgrade hardware or software
    a. when use of system is least critical
    b. replace the old system exactly as it was if problems occur with the new one
    c. before installing a new version of software, the old one should be thoroughly backed up or left in place in another directory so that it can be used if necessary

U. Training of staff and transition techniques

1. effective staff training
    a. have potential operators participate in the design of the system
    b. have operators receive short demonstrations
    c. provide hands-on lessons at the time the system is installed

2. continuing education for staff
    a. national meetings
    b. advanced

V. Unresolved issues



A. What is the balance between confidentiality and free access to clinical records for public health purposes?

B. What is the cost of data access and of programming and processing?

C. What are the methods for unifying data with driver formats?

D. What is the ability of professionals and the public to deal with incomplete and preliminary data?

E. How can features be standardized so that a thoroughly debugged core system can be used?

Return to Lesson 12